

# Identification of effective visual problem solving strategies in a complex visual domain

## Citation for published version (APA):

van Meeuwen, L. W., Jarodzka, H., Brand-Gruwel, S., Kirschner, P. A., de Bock, J. J. P. R., & van Merriënboer, J. J. G. (2014). Identification of effective visual problem solving strategies in a complex visual domain. *Learning and Instruction*, 32, 10-21. <https://doi.org/10.1016/j.learninstruc.2014.01.004>

## Document status and date:

Published: 01/08/2014

## DOI:

[10.1016/j.learninstruc.2014.01.004](https://doi.org/10.1016/j.learninstruc.2014.01.004)

## Document Version:

Publisher's PDF, also known as Version of record

## Document license:

Taverne

## Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

## General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

[www.umlib.nl/taverne-license](http://www.umlib.nl/taverne-license)

## Take down policy

If you believe that this document breaches copyright please contact us at:

[repository@maastrichtuniversity.nl](mailto:repository@maastrichtuniversity.nl)

providing details and we will investigate your claim.



# Identification of effective visual problem solving strategies in a complex visual domain



Ludo W. van Meeuwen<sup>a,b,\*</sup>, Halszka Jarodzka<sup>a</sup>, Saskia Brand-Gruwel<sup>a</sup>, Paul A. Kirschner<sup>a</sup>,  
Jeano J.P.R. de Bock<sup>b</sup>, Jeroen J.G. van Merriënboer<sup>a,c</sup>

<sup>a</sup> Open University of The Netherlands, Heerlen, The Netherlands

<sup>b</sup> Air Traffic Control The Netherlands, Schiphol-Oost, The Netherlands

<sup>c</sup> Maastricht University, Maastricht, The Netherlands

## ARTICLE INFO

### Article history:

Received 13 July 2013

Received in revised form

7 January 2014

Accepted 8 January 2014

Available online 26 January 2014

### Keywords:

Expertise

Eye-tracking

Perceptual task

Visual problem solving

Solution similarity

## ABSTRACT

Students in complex visual domains must acquire visual problem solving strategies that allow them to make fast decisions and come up with good solutions to real-time problems. In this study, 31 air traffic controllers at different levels of expertise (novice, intermediate, expert) were confronted with 9 problem situations depicted on a radar screen. Participants were asked to provide the optimal order of arrival of all depicted aircrafts. Eye-movements, time-on-task, perceived mental effort, and task performance were recorded. Eye-tracking data revealed that novices use inefficient means-end visual problem solving strategies in which they primarily focus on the destination of aircraft. Higher levels of expertise yield visual problem solving strategies characterized by more efficient retrieval of relevant information and more efficient scan paths. Furthermore, experts' solutions were more similar than intermediates' solutions and intermediates' solutions were more similar than novices' solutions. Performance measures showed that experts and intermediates reached better solutions than novices, and that experts were faster and invested less mental effort than intermediates and novices. These findings may help creating eye-movement modeling examples for the teaching of visual problem solving strategies in complex visual domains.

© 2014 Elsevier Ltd. All rights reserved.

## 1. Introduction

In many complex cognitive domains, professionals (e.g., medical specialists, power plant controllers, pilots) make decisions on the basis of their interpretation of complex visualizations. Air traffic controllers, for example, need to interpret available visual information on a radar screen in order to guide aircraft to an airport. Students in air traffic control (ATC) must develop domain-specific visual problem solving strategies to become experts in their domain. Process-oriented worked examples that make the cognitive processes of experts visible can help students learn to solve particular problems (Van Gog, Paas, & Van Merriënboer, 2006, 2008; Van Gog & Rummel, 2010). In visual domains, eye-movements are a direct indicator of visual expertise because they change as experience increases from novice towards expert (for

overviews, see Gegenfurtner, Lehtinen, & Säljö, 2011; Gegenfurtner, Siewiorek, Lehtinen, & Säljö, 2013; Reingold & Sheridan, 2011; Spivey & Dale, 2011). So-called eye-movement modeling examples (EMMEs) may make the visual problem solving process visible by superimposing an expert's gaze pattern on the image so that the learner can study what an expert is looking at and in which order (Jarodzka et al., 2012; Jarodzka, Van Gog, Dorr, Scheiter, & Gerjets, 2013; Van Gog, Jarodzka, Scheiter, Gerjets, & Paas, 2009). However, there are open questions in terms of how to design EMMEs using experts' eye-movements. The first question concerns the strategies for visual problem solving used at different levels of expertise (Feldon, 2007; Jarodzka, Scheiter, Gerjets, & Van Gog, 2010). The second question is whether these strategies lead to one common solution or to a wide variety of solutions when carrying out a perceptual task (cf. Medin et al., 2006).

With regard to strategies used at different levels of expertise, at least three levels can be distinguished in the development towards expert performance (Berliner, 1986; Boshuizen & Schmidt, 2008; Dreyfus & Dreyfus, 2005). Novices are beginners in a domain without relevant experience; they gradually build up a large amount of knowledge which is represented in networks that result

\* Corresponding author. Open University of The Netherlands, Valkenburgerweg 177, 6419 AT Heerlen, The Netherlands. Tel.: +31 45 5762464.

E-mail address: [Ludo.vanMeeuwen@ou.nl](mailto:Ludo.vanMeeuwen@ou.nl) (L.W. van Meeuwen).

in long chains of detailed reasoning steps. Intermediates have already acquired first experiences in a domain, which allows them to encapsulate parts of their knowledge leading to shortcuts in reasoning and thus higher performance. Experts' knowledge is stored in an entirely different, but very efficient manner, namely in scripts (Boshuizen & Schmidt, 2008), enabling them to show "consistently superior performance on a specified set of representative tasks for a domain" (Ericsson & Lehmann, 1996, p. 277). Most research on visual problem solving focused on experts only or on differences between novices and experts. The number of studies using intermediates is limited (Gegenfurtner et al., 2011; Reingold & Sheridan, 2011), and, thus, there is a lack of knowledge about stages in the development of visual problem solving as well as the strategies novices, intermediates and experts use when solving visual problems. This knowledge is needed for designing example-based learning materials such as EMMs. Moreover, it is important to know whether particular visual problem solving strategies lead to different solutions for the same problem or not; obviously, it is more desirable to teach problem solving strategies that lead to similar and good solutions for a wide range of problems.

This article aims at gaining insight in how expertise affects visual problem solving strategies, similarity of found solutions, and performance. The next sections discuss the visual problem solving strategies novices, intermediates and experts use when carrying out perceptual tasks; the degree to which people with different expertise levels and strategies come up with either common or different solutions for the problem at hand, and the moderating effect of task difficulty when studying the influence of expertise on visual problem solving strategies, the similarity of solutions, and performance.

### 1.1. Visual problem solving strategies

When solving problems, cognitive schemas retrieved from long-term memory enable the use of problem solving strategies (Boshuizen & Schmidt, 2008). At least three problem solving strategies can be distinguished for solving visual problems, namely, attention focusing, chunking, and means-end analysis (Chi, Glaser, & Rees, 1982; Gobet & Simon, 1998; Haider & Frensch, 1999; Simon, 1975).

When using the strategy *attention focusing* schemas help to distinguish between relevant and irrelevant information and so enable problem solvers to focus on what is important in a given problem situation. Haider and Frensch (1999) describe in their *information-reduction* theory that experts optimize the amount of processed information by separating task-irrelevant from task-relevant information. This theory was supported by the findings in a meta-analysis by Gegenfurtner et al. (2011) and by Reingold and Sheridan's (2011) review of research on expertise in medicine and chess. In the field of aviation, two studies support the information-reduction theory. Kasarskis, Stehwien, Kickox, and Aretz (2001) studied scanning characteristics of novice and expert aircraft pilots during landing. They found that eye-scanning patterns and specific fixation behaviors of experts differed from those of novices. Experts showed shorter but relatively more eye-fixations (during fixations the eyes stand still and take in new information), more eye-fixations on relevant points such as aim point and airspeed, and fewer fixations on less relevant points such as the altimeter because all necessary altitude information was obtainable from the true horizon. Also in a study by Bellenkes, Wickens, and Kramer (1997), expert pilots scanned more crucial instruments during a simulated flight task than novices. In ATC the use of information reduction could show in experts fixating faster on relevant objects (i.e., aircraft) and fixating them relatively longer.

The strategy of *perceptual chunking* relevant information, described as *unitization* by Goldstone (1998), makes it possible to

combine important elements together so that they can be treated in working memory as *one* information element in a given problem situation. This requires less effort than processing all elements separately. For example, experts in chess are known to become familiar with complex configurations of separate chessmen and they are able to recognize these configurations as single units (Jongman, 1968). Hence, experts use schemas formed from earlier experiences and recognize familiar compositions of task elements or 'patterns' (e.g., frequently occurring air traffic situations) without viewing all the details (Gobet & Simon, 1998). In ATC, the use of perceptual chunking would be manifest in less gaze switches (i.e., transitions) between separate elements (e.g., aircraft), because particular groups of elements (e.g., all aircraft in a queue) are perceived as one element (i.e., chunk).

The strategy that can be characterized as *means-end analysis* is based on schemas for working backward from the goal, rather than working towards the goal. This strategy is described as a highly general but effort-demanding problem solving strategy (Simon, 1975), where the task performer uses a continuous orientation on the goal (the 'end') and tentatively applies operators (the 'means') to determine a next step in the problem solving process that helps to move in the direction of the goal. More advanced problem solvers understand which routine of operations is underlying the final solution. Thus, they do not reason backwards from the goal but decide based on the prior act what the next act should be to reach the final goal. This sequence of actions can ultimately become automated, leading to fast and accurate performance which hardly requires the investment of mental effort (Chi et al., 1982; Sweller, 2004; Van Merriënboer, Clark, & De Croock, 2002). In a visual domain like ATC, the use of means-end analysis would be manifest by frequently focusing (i.e., more eye-fixations) on the goal (e.g., the airport), whereas working-forward strategies would be manifest by frequently focusing on the elements that are affected by the problem solving steps (e.g., the aircraft).

### 1.2. Similarity of solutions

For problems in complex visual domains, there is typically not one general problem solution but a broad range of solutions that may vary from suboptimal (or even incorrect) to more optimal (Gronlund, Dougherty, Durso, Canning, & Mills, 2005; Mumford, Schultz, & Van Doorn, 2001). In ATC, for example, the number of acceptable solutions to guide the aircraft to an airport is restricted by safety rules and the need for efficiency (safety: maintaining at least five miles horizontal separation and 1000 feet vertical separation; efficiency: causing as little delays as possible), but there are many degrees of freedom in finding these solutions (e.g., you can keep enough separation between aircraft by changing either their speed, height or direction).

The level of expertise influences the ability of anticipating on possible situations (Mumford et al., 2001) resulting in more or less optimal solutions. For novices, visual problem solving is highly demanding because they have not yet cognitive schemas available that help them organize the perceived information. Due to their limited working-memory capacity they are easily overwhelmed by the amount of information, especially when this information is transient such as in ATC (Lowe, 2003; Mayer, 2005; Scheiter, Gerjets, Huk, Imhof, & Kammerer, 2009; Spanjers, Van Gog, & Van Merriënboer, 2010; Sweller, Van Merriënboer, & Paas, 1998). As a consequence, their awareness of the current situation will be limited, incomplete and sometimes erroneous, which hampers their projection of the future status (Endsley, 1995) and thus leads to a broad range of dissimilar solutions, including many incorrect or suboptimal solutions.

Intermediates have constructed cognitive schemas that allow them to organize given information when they are confronted with visual problems, but, compared to experts they have more problems with linking their schemas to specific problem situations (Boshuizen & Schmidt, 1992, 2008). They still have difficulties with immediate pattern recognition or may be not aware that a chosen schema is not appropriate for the given problem and thus miss important details for correct decision making (for an overview, see Gegenfurtner et al., 2013). It can thus be expected that intermediates find better and more similar solutions than novices, but there will still be a notable dissimilarity across solutions because they frequently come up with less optimal solutions.

Experts possess well developed schemas for many specific situations (e.g., in ATC: “inbound traffic from the west with strong tail wind”), which help them to quickly build a good awareness of the problem situation (Boshuizen & Schmidt, 2008; Schmidt, Norman, & Boshuizen, 1990). The number of available schemas also increases experts’ repertoire for solving problems (De Groot, 1978; Ericsson & Lehmann, 1996; Gobet & Simon, 1998), resulting in a flexible range of potential problem solving strategies (Lesgold et al., 1988). Their ability to quickly recognize a broad range of problem situations allows them to come up with optimal solutions (Endsley, 2006), and because most of these solutions are optimal they can be expected to be relatively similar. If any differences occur, these will be marginal and based on personal preferences.

### 1.3. Performance and task difficulty

A better understanding of visual problem solving strategies and similarity of solutions will help to explain differences in performance between experts, intermediates and novices. Such performance differences have been well documented in literature, showing that experts outperform intermediates and novices, and intermediates outperform novices. First, higher expertise is associated with higher accuracy and reaching more optimal solutions (Ericsson, 2006). Second, higher expertise is associated with faster performance or speed, meaning that experts not only reach better solutions but also do this in less time (Lesgold et al., 1988). Third, experts have better developed schemas allowing them to make changes already early which means that they prevent conflicts later in the process, resulting in lower mental effort during task performance (Sweller et al., 1998).

Performance differences between experts, intermediates and novices may not show for all levels of task difficulty (in ATC: amount of aircraft that must be controlled, potential conflicts, weather conditions). For example, novices and intermediates may perform equally well on very easy tasks, while intermediates and experts may perform equally well on tasks at a medium level of difficulty. Thus, it is important to compare expertise levels across tasks with different difficulty levels, and performance differences between experts, intermediates, and novices may become more visible as tasks become more difficult. Furthermore, eye-movements and visual problem solving strategies may also vary as a function of task difficulty (cf. Gegenfurtner et al., 2011). Less experienced problem solvers are not yet able to ignore irrelevant information, chunk related elements and work forward, which causes conflicts with the limited capacity of processing capacity available, especially for more difficult tasks (Mayer, 2005; Mayer & Moreno, 2003). Therefore, differences in visual problem solving strategies and similarities of solutions between experts, intermediates and novices are also expected to become more visible in more difficult tasks.

### 1.4. Hypotheses

Experts, intermediates and novices are expected to use different visual problem solving strategies and will thus show different eye-

movements (Hypothesis 1). First, experts will have a better information-reduction strategy resulting in more eye-fixations on relevant areas of interest and shorter times to the first fixation on these relevant areas of interest compared to intermediates and novices (Hypothesis 1a). Second, experts will have a better perceptual chunking strategy resulting in less gaze switches between single elements (e.g., they deal with groups of aircraft rather than a single aircraft) and thus more efficient scan paths compared to intermediates and novices (Hypothesis 1b). Third, experts will not use means-end analysis but a working-forward strategy, resulting in less eye-fixations on the destination point of the aircraft compared to intermediates and novices (Hypothesis 1c).

Furthermore, experts are expected to reach more similar task solutions than intermediates, and both experts and intermediates are expected to reach more similar solutions than novices (Hypothesis 2).

For the quality of performance, experts are expected to reach better solutions in less time and to invest less mental effort than intermediates, and both experts and intermediates are expected to reach better solutions in less time and to invest less mental effort than novices (Hypothesis 3).

Finally, the differences between experts, intermediates and novices are expected to be more pronounced for relatively difficult tasks than for easy tasks, yielding interactions between level of expertise and task difficulty (Hypothesis 4).

## 2. Method

### 2.1. Participants

Participants in the study were 31 individuals (age:  $M = 26.45$  years,  $SD = 6.31$ ; 8 females and 23 males) with three different levels of expertise. All participants were employed at Air Traffic Control, The Netherlands. Experts were ten fully licensed air traffic controllers whom had worked for at least two years (years of work experience:  $M = 7.10$ ,  $SD = 6.83$ ; age:  $M = 33.10$  years,  $SD = 6.81$ ; 2 females and 8 males). Intermediates were nine students in the final phase of the regular on-the-job-training program for air traffic controllers (months of training:  $M = 22.33$ ,  $SD = 6.20$ ; age:  $M = 24.67$  years,  $SD = 2.18$ ; 3 females and 6 males). Novices were 12 students in the initial phase of the ATC training program (months of training:  $M = 3.25$ ,  $SD = 0.45$ ; age:  $M = 22.25$ ,  $SD = 2.30$ ; 3 females and 9 males).

### 2.2. Materials and apparatus

#### 2.2.1. Air-traffic control tasks

Nine tasks with three difficulty levels (i.e., three easy, three medium, and three difficult tasks) were composed using still pictures of realistic ATC radar situations. The three levels of difficulty were determined a priori, depending on the number of aircraft involved and the number of conflicts ahead. The tasks were composed by a domain expert in ATC and involved a number of inbound aircraft heading towards the initial approach fix “Artip” of Amsterdam Airport Schiphol (see Fig. 1). Artip is the route junction and initial approach fix that aircraft need to cross (i.e., the ‘goal’). For each task, participants had to determine the optimal order of arrival at Artip of the aircraft that had to be controlled (e.g., KLM078, KLM1234, TRA321, JAP411 etc.). Three sets of tasks were composed and each set comprised three of the nine tasks. A set started with an easy task, followed by a medium task, and finally a difficult task. To avert any order effects, the order of the sets was counterbalanced between participants.

ATC is a dynamic task but there are reasons to use still pictures in this study. Stills of ATC radar situations were used to create

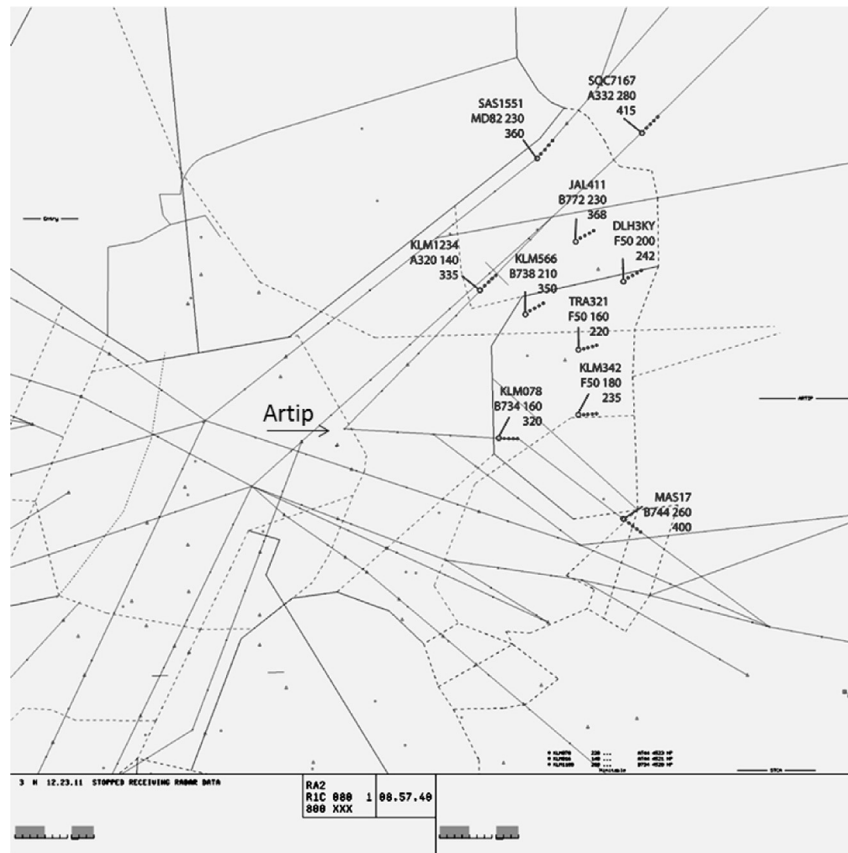


Fig. 1. Negative of example screenshot of an ATC radar screen. The ATC controller has to determine the optimal order of arrival of the aircraft to the initial approach fix Artip.

constrained processing tasks (cf. Hoffman, 1987), suitable for conducting analyses on the initial phase of the control process. In ATC, in the initial phase situational awareness is built to decide on the optimal order of arrival of the aircraft (Oprins & Schuver, 2003). Moreover, the analysis of stills allows for the use of fixation parameters in eye-tracking data, which would not be possible in dynamic stimuli due to the occurrence of smooth pursuit. Apart from time pressure, the use of stills allows to capture all relevant processes.

### 2.2.2. Mental effort

For each task the perceived mental effort was measured using the scale developed by Paas (1992). Participants indicated their perceived mental effort after accomplishing the task on a 9-point Likert scale ranging from 1 ("very, very low effort") to 9 ("very, very high effort").

### 2.2.3. Eye-tracking

Because the use of strategies is based on visual information, eye-tracking is a plausible technique to provide evidence on the strategies used when carrying out perceptual tasks. Eye-tracking research distinguishes between fixations and saccades (Holmqvist et al., 2011). During fixations the eyes stand still and can take in new information. Saccades are the eye-movements from one fixation to another. During saccades, the information transfer is suppressed but the focus of attention moves from one element to another element. Saccades between several areas are also referred to as *transitions*.

The still ATC pictures were presented on a 17" (43 cm) diagonal screen (1280\*1024 pixels). During task performance, eye-

movements of participants were recorded with a Tobii 1750 remote eye-tracking system with a temporal resolution of 50 Hz. Eye-movement data were recorded and processed with Tobii Studio 2.1 software using the standard Tobii fixation filter algorithm (settings: 35 velocity × 35 dispersion).

### 2.3. Procedure

The experiment was run in individual sessions of approximately 60 min. First the participants answered some demographic questions. Subsequently, participants received the following instructions: "You will see nine radar-ATC situations that are presented one by one. For each situation, please execute the following task: Determine the preferred order of arrival at point Artip and report this order out loud. Work as safe, as efficient, and as fast as possible". After the warming-up task in which Artip was indicated, the nine still pictures of ATC radar situations were presented to the participants one by one. After completing each of the nine tasks, participants indicated their perceived mental effort.

### 2.4. Data analysis

An overview of all dependent measures can be found in Table 1.

#### 2.4.1. Analysis of visual strategies

To assign eye-tracking data to an element or region on the screen, each still ATC picture was divided into areas of interest (AOI). Visual inspection of the eye-tracking data yielded insight in the minimal size of the AOIs. Each AOI initially had the same size and either covered the radar plot, the trail indicated with dots, the



**Table 1**  
Descriptions and explanations of all dependent variables.

	Calculation	Explanation
<i>Eye-tracking measures</i>		
Relative Fixation Duration on Aircraft	Dwell on aircraft AOI (% of total time-on-task)	Time participants spent looking at the different aircraft
Relative Fixation Duration on Target	Dwell on Artip AOI (% of total time-on-task)	Time participants spent looking at the aim point of the aircraft to arrive at the airport
Relative fixation Duration on Background area	Dwell on background (% of total time-on-task)	Time participants spent looking at the background area (i.e., not looking at any AOI but looking elsewhere on the radar plot)
Time to First Fixations on Aircraft	Average time until the participant looked at aircraft on the screen for the first time from stimulus onset	Time elapsed until participants first looked at any aircraft.
Time to First Fixations on Artip	Average time until the participant looked at the target area on the screen (i.e., Artip) for the first time from stimulus onset	Time elapsed until participants first looked at the aim point of the aircraft to arrive at the airport
Number of Fixations on Aircraft	Sum of fixations on aircraft AOI	Number of looks (i.e., fixations) on any aircraft
Number of Fixations on Artip	Sum of fixations on Artip AOI	Number of looks (i.e., fixations) at the aim point of the aircraft
Transitions Artip – Aircraft	Number of participants' fixation switches from an aircraft AOI to the target AOI and vice versa	Number of switching between looking at one of the aircraft and looking at the aim point of the airport
Transitions Aircraft (X) – Aircraft (Y)	Number of participants' fixation switches from an aircraft AOI to another aircraft AOI	Number of switching between looking at one aircraft and looking at another aircraft
Transitions Artip – Background	Number of participants' fixation switches from the target AOI to the background AOI and vice versa	Number of switching between looking at one of the aim point and looking at the background area
Transitions Background – Aircraft	Number of participants' fixation switches from an aircraft AOI to the background AOI and vice versa	Number of switching between looking at the background area and looking at one of the aircrafts
<i>Similarity measure</i>		
Similarity of Task Solutions	Levenshtein distance between two task solutions	Participants are asked to provide an order in which the aircraft arrive in the landing point, resulting in a string of aircraft names. We compared each participant's order to the rest of his or her group. This was done by transforming one aircraft string to the other (either by insertion, deletion, or substitution), calculating the number of transitions made, and transforming this into a percentage value.
<i>Performance measures</i>		
Task Correctness Score	Maximum score (i.e., 5 points) minus one point per unrealistic order	Participants' score on their proposed order of arrival of aircraft
Time-on-task	Time between stimulus onset until a solution is given	Participants' time to decide the order of arrival of all aircraft
Perceived Mental Effort	Participants' perceived mental effort per task	Subjectively rated amount of cognitive capacity used for a task

label of a single aircraft, or Artip. In situations where aircraft were close to each other, the related AOIs overlapped. In such cases, the overlapping AOIs were equally reduced in size until they were exactly adjacent to each other (for an example, see Fig. 2).

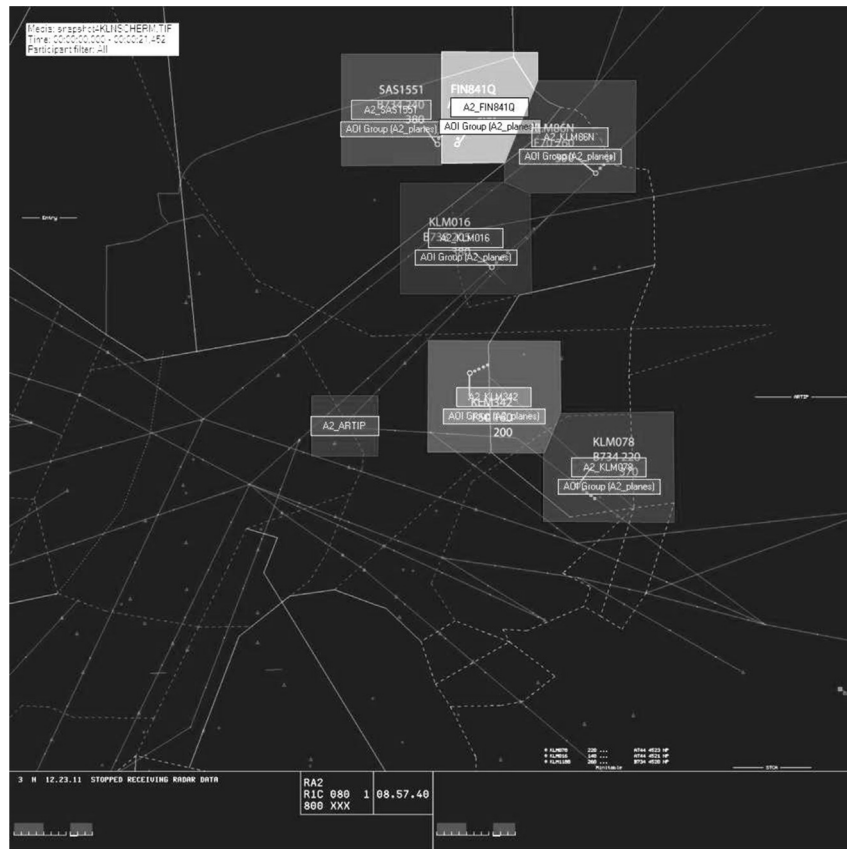
The following eye-tracking measures were derived per AOI: Relative fixation duration (i.e., total time spent looking at a certain area on the screen), time until first fixation (i.e., the time until the participant looked at a certain area on the screen for the first time from stimulus onset), the number of fixations (i.e., number of looks at a certain area on the screen), and AOI transition matrices (i.e., indicators for how often participants switched their gaze from one AOI to another AOI). Because time-on-task differed across participants, also the total fixation duration did. Hence, to make this fixation duration comparable across participants and across tasks, the sum of total fixation durations on AOIs and non-AOI areas was standardized by dividing each sum by the individual time spent on task. This resulted in *relative fixation duration* measures which indicated the ratio of participants' attention. Missing values (i.e., no visit on AOI) were replaced by zero. In case of no visit on AOI, no time to first fixation was recorded either. Hence, to make this eye-tracking parameter comparable across participants, missing values were replaced by the maximum time-on-task across participants (see for a description of the same procedure, Jarodzka et al., 2010). To obtain transition matrices, individual strings of all fixation locations were exported from Tobii Studio and transformed per ATC task into matrices. These matrices comprised per task the number

of transitions for each participant between and within all different AOIs. Per task and per participant the total number of transitions between different aircraft and the total number of transitions within aircraft were computed. Per task only one AOI covered Artip and the background area was defined as the area not being covered by an AOI (i.e., non-AOI area). Finally, the total number of transitions between aircraft and background, aircraft and Artip, and background and Artip were computed.

#### 2.4.2. Analysis of similarity of task solutions

Two experts blindly and independently scored all performances on one of the difficult tasks (i.e., more than 10% of the tasks). They subtracted one point from the maximum of five points for each unrealistic order of aircraft resulting in a *task correctness* score. The maximum number of points subtracted was five so that the scores ranged between 0 and 5. They achieved a high inter-rater reliability: Spearman rank correlation  $\rho = .846$  ( $p < .001$ ). The remaining tasks were scored by only one of the experts.

The similarity of task solutions was calculated by means of sequence analyses based on the so-called Levenshtein distances (Levenshtein, 1966). The Levenshtein distance is a measure for difference between two sequences. It is obtained by the minimal number of operations needed to transfer one sequence into another sequence. The possible operations are insertions, deletions, or substitutions of single characters. To determine the Levenshtein distance in the present study, a string of aircraft that a participant



**Fig. 2.** Examples of Areas of Interest (AOIs) around aircraft and Artip. Size of the AOIs is determined by the average size and shape of an eye-tracking heat map. The AOIs around the aircraft were adjusted in size to not overlap each other.

determined in a given order (e.g., KLM078, KLM1234, TRA321, JAP411, ...) served as the input data. The number of operations (i.e., insertions or deletions of aircraft) needed to transform the aircraft sequence of this participant into that of another participant describes the difference between the two sequences of aircraft. The Levenshtein distance was determined for the aircraft sequences of experts, intermediates and novices to analyze the similarity of task solutions used within these groups. This procedure resulted in a similarity score for each possible pair of experts, intermediates, and novices.

For all analyses a factorial repeated-measures ANOVA is used with levels of expertise (i.e., novice, intermediate, expert) as between-subjects factor and difficulty (i.e., easy, medium, and difficult) as within-subjects factor. A significance level of .05 is used for all reported analyses. To test the hypotheses, the main effects of expertise level and task difficulty and their interaction are reported. In case of a significant effect, Bonferroni post-hoc tests are conducted. Because of problems with sphericity, the results of the Greenhouse Geisser contrast analysis are given.

### 3. Results

#### 3.1. Visual problem solving strategies

Means and standard deviations of all eye-tracking measures are presented in Table 2.

##### 3.1.1. Fixation duration

To test Hypotheses 1a (i.e., information reduction strategy) and 1b (i.e., perceptual chunking strategy), the *relative fixation duration on aircraft*, *relative fixation duration on Artip*, and *relative fixation*

*duration on the background area* were analyzed. Relative fixation durations are given as percentage of time-on-task.

Results showed no main effect of expertise level on relative fixation duration on aircraft,  $F(2, 29) = 1.240$ ,  $MSE = 16.146$ ,  $p = .305$ ,  $\eta_p^2 = .081$ , but it showed a main effect of task difficulty,  $F(2, 58) = 32.02$ ,  $MSE = 36.73$ ,  $p < .001$ ,  $\eta_p^2 = .534$ . Post-hoc tests revealed that relative fixation duration on aircraft in easy tasks was significantly shorter than in medium tasks ( $p = .015$ ), and that relative fixation duration on aircraft in both easy tasks and medium tasks was shorter than in difficult tasks (both  $p$ -values  $< .001$ ). No interaction effect was found between expertise level and task difficulty on relative fixation duration on aircraft,  $F(4, 58) = 2.45$ ,  $MSE = 36.37$ ,  $p = .060$ ,  $\eta_p^2 = .149$ .

Results showed a main effect of expertise level on relative fixation duration on Artip,  $F(2, 29) = 8.96$ ,  $MSE = .19$ ,  $p = .001$ ,  $\eta_p^2 = .39$ , as well as a main effect of task difficulty,  $F(2, 58) = 6.88$ ,  $MSE = .617$ ,  $p = .008$ ,  $\eta_p^2 = .197$ . Post-hoc tests in expertise levels revealed that novices fixated longer on Artip than both intermediates ( $p = .009$ ) and experts ( $p = .002$ ). Post-hoc tests in task difficulty revealed that relative fixation duration on Artip in easy tasks was longer than in difficult tasks ( $p = .024$ ). Moreover, results showed an interaction effect between expertise level and task difficulty,  $F(4, 58) = 4.174$ ,  $MSE = .617$ ,  $p = .015$ ,  $\eta_p^2 = .230$  (see Fig. 3a). Post-hoc tests revealed that on easy tasks novices fixated longer on Artip than both experts ( $p = .001$ ) and intermediates ( $p = .002$ ); on medium tasks, novices fixated longer on Artip than experts ( $p = .015$ ), and on difficult tasks novices fixated longer on Artip than both experts ( $p = .010$ ) and intermediates ( $p = .010$ ).

Results showed a main effect of expertise level on relative fixation duration on the background area,  $F(2, 29) = 4.03$ ,  $MSE = 19.94$ ,  $p = .029$ ,  $\eta_p^2 = .224$ , as well as a main effect of task difficulty,  $F(2,$

**Table 2**

Means and standard deviations of all dependent variables.

	Experts			Intermediates			Novices		
	Easy <i>M</i> ( <i>SD</i> )	Medium <i>M</i> ( <i>SD</i> )	Difficult <i>M</i> ( <i>SD</i> )	Easy <i>M</i> ( <i>SD</i> )	Medium <i>M</i> ( <i>SD</i> )	Difficult <i>M</i> ( <i>SD</i> )	Easy <i>M</i> ( <i>SD</i> )	Medium <i>M</i> ( <i>SD</i> )	Difficult <i>M</i> ( <i>SD</i> )
<i>Eye-tracking</i>									
Relative Fixation Duration Aircraft (% Total time-on-task)	88.86 (4.18)	89.81 (5.31)	94.27 (3.50)	86.01 (6.73)	86.77 (8.89)	91.87 (4.28)	85.17 (4.13)	91.40 (3.21)	94.65 (2.00)
Relative Fixation Duration on Artip (% Total time-on-task)	0.06 (0.14)	0.01 (0.04)	0.06 (0.15)	0.23 (0.49)	0.16 (0.27)	0.04 (0.08)	1.25 (1.09)	0.64 (0.71)	0.43 (0.45)
Relative fixation Duration on Background area (% Total time-on-task)	3.27 (3.17)	5.33 (5.94)	2.49 (2.06)	10.40 (7.36)	11.96 (10.26)	6.19 (4.30)	10.20 (3.89)	5.61 (3.05)	2.75 (1.23)
Time to First Fixations on Aircraft (Sum of Time to First Fixations on Planes in seconds)	22.65 (6.01)	21.30 (7.23)	80.82 (22.57)	24.02 (6.58)	32.22 (6.55)	107.56 (21.43)	30.02 (8.97)	34.65 (9.51)	130.50 (39.24)
Time to First Fixations on Artip (in seconds)	30.07 (3.01)	40.90 (4.00)	44.18 (6.18)	28.70 (6.16)	34.39 (12.56)	46.78 (.26)	18.28 (10.24)	29.16 (13.68)	28.68 (19.83)
Number of Fixations on Aircraft	13.87 (4.91)	27.83 (10.26)	62.13 (16.65)	18.96 (5.26)	35.00 (8.39)	78.11 (18.37)	22.94 (6.57)	42.25 (14.06)	77.31 (21.50)
Number of Fixations on Artip	0.07 (0.14)	0.03 (0.11)	0.07 (0.14)	0.11 (0.24)	0.19 (0.29)	0.07 (0.15)	0.72 (0.57)	0.58 (0.61)	0.64 (0.67)
Transitions Artip – Planes	0.10 (0.16)	0.03 (0.11)	0.07 (0.21)	0.260 (0.32)	0.11 (0.17)	0.70 (1.42)	0.81 (0.70)	0.72 (0.68)	1.36 (1.66)
Transitions Aircraft (X) – Aircraft (Y)	8.83 (3.32)	17.90 (7.99)	46.13 (14.70)	11.00 (3.17)	20.93 (6.34)	53.96 (13.46)	13.28 (3.83)	27.140 (9.70)	55.50 (18.33)
Transitions Artip – Background	0.03 (0.11)	0.00 (0.00)	0.13 (0.28)	0.07 (0.22)	0.11 (0.24)	0.07 (0.15)	0.47 (0.50)	0.33 (0.47)	0.22 (0.26)
Transitions Background – Aircraft	1.37 (0.90)	3.30 (2.43)	4.27 (3.16)	4.70 (2.93)	9.93 (6.99)	9.52 (5.12)	6.67 (3.44)	7.31 (3.99)	6.32 (3.79)
<i>Similarity</i>									
Similarity of Task Solutions	.96 (.09)	.59 (.17)	.23 (.07)	.91 (.11)	.47 (.19)	.21 (.05)	.71 (.19)	.40 (.16)	.18 (.04)
<i>Performance</i>									
Task Correctness Score	4.97 (0.11)	4.83 (0.28)	4.08 (0.79)	4.78 (0.55)	4.48 (0.63)	3.65 (0.79)	4.53 (0.61)	3.97 (0.56)	2.97 (0.942)
Time-on-task	9.51 (2.36)	16.28 (2.98)	31.19 (5.65)	12.26 (2.78)	21.85 (4.06)	40.28 (6.70)	16.23 (4.45)	25.79 (6.55)	44.76 (8.71)
Perceived Mental Effort	2.00 (0.70)	3.80 (1.06)	5.30 (1.44)	3.41 (1.01)	4.78 (0.76)	6.07 (0.85)	3.19 (0.89)	4.39 (0.57)	5.56 (0.50)

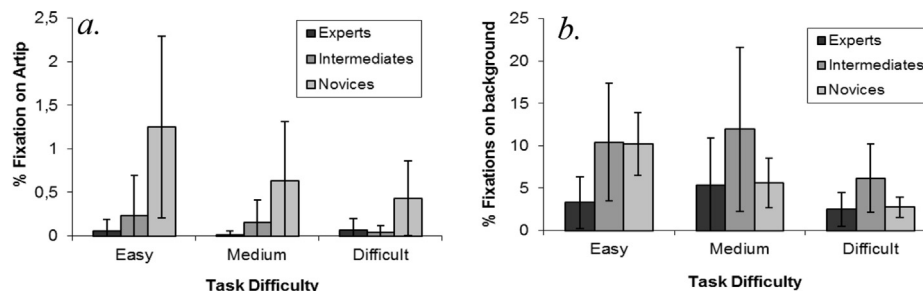
58) = 30.51,  $MSE = 122.32$ ,  $p < .001$ ,  $\eta_p^2 = .521$ . Post-hoc tests between expertise levels revealed that intermediates fixated longer on the background area than experts ( $p = .025$ ), while novices did not differ from intermediates ( $p = .307$ ) and experts ( $p = .609$ ). Post-hoc tests between task difficulty levels revealed that relative fixation duration on the background area in easy tasks was significantly longer than in both medium and difficult tasks (both  $p$ -values  $< .001$ ). Moreover, results showed an interaction effect between expertise level and task difficulty,  $F(4, 58) = 5.835$ ,  $MSE = 28.791$ ,  $p = .001$ ,  $\eta_p^2 = .294$  (see Fig. 3b). Post-hoc tests revealed that on easy tasks experts fixated less on the background area than both intermediates ( $p = .012$ ) and novices ( $p = .009$ ). On difficult tasks, intermediates fixated more on the background area than both experts ( $p = .017$ ) and novices ( $p = .022$ ).

### 3.1.2. Time to first fixation

To further test Hypothesis 1a (i.e., information reduction strategy), the time to first fixation on aircraft and time to first fixation on Artip were analyzed.

Results showed a main effect of expertise level on time to first fixation on aircraft,  $F(2, 29) = 8.21$ ,  $MSE = 158.44$ ,  $p = .002$ ,  $\eta_p^2 = .37$ , as well as a main effect of task difficulty,  $F(2, 58) = 327.32$ ,  $MSE = 1402.47$ ,  $p < .001$ ,  $\eta_p^2 = .921$ . Post-hoc tests between expertise levels revealed that novices took significantly more time than experts to first fixate on aircraft ( $p = .001$ ), while intermediates did not differ from novices ( $p = .334$ ) and experts ( $p = .109$ ). Post-hoc tests between task difficulty levels revealed that time to first fixation on aircraft in difficult tasks was longer than in both medium tasks and easy tasks, and longer in medium tasks than in easy tasks (all  $p$ -values  $< .001$ ). Moreover, results showed an interaction effect between expertise level and task difficulty,  $F(4, 58) = 6.813$ ,  $MSE = 1402.47$ ,  $p = .003$ ,  $\eta_p^2 = .327$  (see Fig. 4). Post-hoc tests revealed that in medium tasks, experts showed a shorter time to first fixation on aircraft than both intermediates ( $p = .019$ ) and novices ( $p = .002$ ), and in difficult tasks, experts showed a shorter time than novices ( $p = .002$ ).

Results also showed a main effect of expertise level on time to first fixation on Artip,  $F(2, 29) = 8.00$ ,  $MSE = 69.02$ ,  $p = .002$ ,



**Fig. 3.** a. Means and standard deviations of fixation duration on Artip by experts, intermediates and novices in easy, medium and difficult tasks. b. Means and standard deviations of fixation duration on the background area by experts, intermediates and novices in easy, medium and difficult tasks.



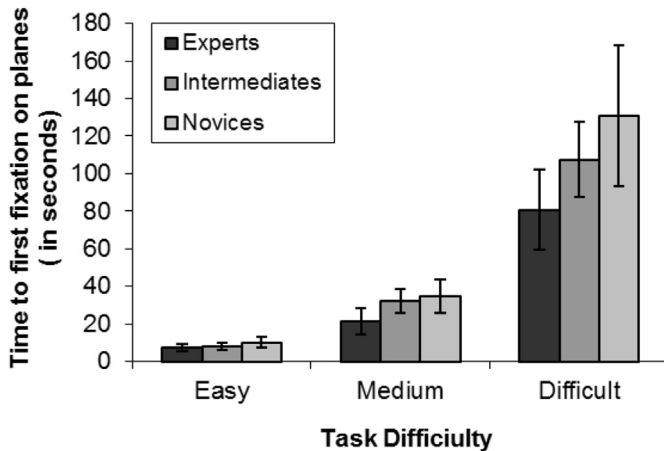


Fig. 4. Means and standard deviations of time to first fixation on aircraft by experts, intermediates and novices in easy, medium, and difficult tasks.

$\eta_p^2 = .364$ , as well as a main effect of task difficulty,  $F(2, 58) = 22.852$ ,  $MSE = 248.82$ ,  $p < .001$ ,  $\eta_p^2 = .449$ . Post-hoc tests between expertise levels revealed that novices took less time to first fixate on Artip than both intermediates ( $p = .014$ ) and experts ( $p = .003$ ), while intermediates and experts did not differ from each other ( $p = 1.00$ ). Post-hoc tests between task difficulty levels revealed that time to first fixation on Artip in easy tasks is shorter than in medium tasks and difficult tasks (both  $p$ -values  $< .001$ ). No interaction effect was found between expertise level and task difficulty,  $F(4, 58) = 1.623$ ,  $MSE = 46.75$ ,  $p = .192$ ,  $\eta_p^2 = .104$ .

### 3.1.3. Number of fixations

To test Hypothesis 1c (i.e., means-end analysis) and to further test Hypothesis 1a, the number of fixations on aircraft and then number of fixations on Artip were analyzed.

Results showed a main effect of expertise level on number of fixations on aircraft,  $F(2, 29) = 3.58$ ,  $MSE = 131.57$ ,  $p = .041$ ,  $\eta_p^2 = .204$ , as well as a main effect of task difficulty,  $F(2, 58) = 341.59$ ,  $MSE = 232.00$ ,  $p < .001$ ,  $\eta_p^2 = .924$ . Post-hoc tests between expertise levels revealed that novices used more fixations on aircraft than experts ( $p = .042$ ). Post-hoc tests between task difficulty levels revealed that number of fixations on aircraft in easy tasks is lower than in both medium tasks and difficult tasks (both  $p$ -values  $< .001$ ) and in medium tasks the number of fixations on aircraft is lower than in difficult tasks ( $p < .001$ ). No interaction effect was found between expertise level and task difficulty,  $F(4, 58) = 1.440$ ,  $MSE = 232.00$ ,  $p = .250$ ,  $\eta_p^2 = .093$ .

Results also showed a main effect of expertise level on number of fixations on Artip,  $F(2, 29) = 8.29$ ,  $MSE = .141$ ,  $p = .001$ ,  $\eta_p^2 = .372$ . No main effect of task difficulty was found,  $F(2, 58) = .314$ ,  $MSE = .143$ ,  $p < .716$ ,  $\eta_p^2 = .011$ . Post-hoc tests between expertise levels revealed that novices used more fixations on Artip than both intermediates ( $p = .011$ ) and experts ( $p = .003$ ). No interaction effect was found between expertise level and task difficulty,  $F(4, 58) = .812$ ,  $MSE = .143$ ,  $p = .516$ ,  $\eta_p^2 = .055$ .

### 3.1.4. Transitions

To further test Hypothesis 1b (i.e., perceptual chunking strategy) and Hypothesis 1c, the number of transitions between Artip and aircraft, different aircraft (e.g., Aircraft X – Aircraft Y), Artip and background (e.g., Artip – some white space around Artip or around the aircraft), and background and aircraft were analyzed.

Results showed a main effect of expertise level on number of transitions between Artip and aircraft,  $F(2, 29) = 6.78$ ,  $MSE = .34$ ,

$p = .004$ ,  $\eta_p^2 = .33$ , but no main effect of task difficulty,  $F(2, 58) = 3.396$ ,  $MSE = 2.085$ ,  $p = .065$ ,  $\eta_p^2 = .108$ . Post-hoc tests between expertise levels revealed that experts used less transitions between Artip and aircraft than novices ( $p = .004$ ), while experts and intermediates ( $p = .861$ ) as well as novices and intermediates ( $p = .079$ ) did not differ significantly from each other. No interaction effect was found,  $F(4, 58) = 1.768$ ,  $MSE = .174$ ,  $p = .149$ ,  $\eta_p^2 = .112$ .

Results showed no main effect of expertise level on number of transitions between different aircraft,  $F(2, 29) = 2.06$ ,  $MSE = 78.07$ ,  $p = .146$ ,  $\eta_p^2 = .13$ , but showed a main effect of task difficulty,  $F(2, 58) = 272.73$ ,  $MSE = 254.81$ ,  $p < .001$ ,  $\eta_p^2 = .907$ . Post-hoc tests between task difficulty levels revealed that the number of transitions between different aircraft in easy tasks was less than in both medium and difficult tasks, and also less in medium tasks than in difficult tasks (all  $p$ -values  $< .001$ ). No interaction effect was found,  $F(4, 58) = .794$ ,  $MSE = 254.81$ ,  $p = .479$ ,  $\eta_p^2 = .054$ .

Results showed a main effect of expertise level on number of transitions between Artip and background,  $F(2, 29) = 5.08$ ,  $MSE = .054$ ,  $p = .013$ ,  $\eta_p^2 = .27$ , but no main effect of task difficulty,  $F(2, 58) = .403$ ,  $MSE = .174$ ,  $p = .668$ ,  $\eta_p^2 = .014$ . Post-hoc tests between expertise levels revealed that experts showed fewer transitions between Artip and background than novices ( $p = .023$ ). Intermediates and novices ( $p = .056$ ) and experts and intermediates ( $p = 1.00$ ) did not differ significantly. No interaction effect was found,  $F(4, 58) = 1.77$ ,  $MSE = .174$ ,  $p = .149$ ,  $\eta_p^2 = .112$ .

Finally, results showed a main effect of expertise level on number of transitions between background and aircraft,  $F(2, 29) = 6.71$ ,  $MSE = 10.13$ ,  $p = .004$ ,  $\eta_p^2 = .32$ , as well as a main effect of task difficulty,  $F(2, 58) = 8.430$ ,  $MSE = 25.546$ ,  $p = .001$ ,  $\eta_p^2 = .231$ . Post-hoc tests between expertise levels revealed that experts used fewer transitions between background and aircraft than both intermediates ( $p = .005$ ) and novices ( $p = .029$ ), while intermediates and novices did not differ significantly from each other ( $p = 1.00$ ). Post-hoc tests between task difficulty levels revealed that the number of transitions between background and aircraft in easy tasks was smaller than in both medium tasks and difficult tasks (both  $p$ -values  $= .002$ ). Moreover, an interaction effect was found between expertise level and task difficulty,  $F(4, 58) = 29.98$ ,  $MSE = 25.545$ ,  $p = .031$ ,  $\eta_p^2 = .176$  (see Fig. 5). Post-hoc tests revealed that on easy tasks experts showed less transitions between background and aircraft than both intermediates ( $p = .037$ ) and novices ( $p < .001$ ). Experts also showed less transitions than intermediates in medium tasks ( $p = .014$ ) and difficult tasks ( $p = .026$ ).

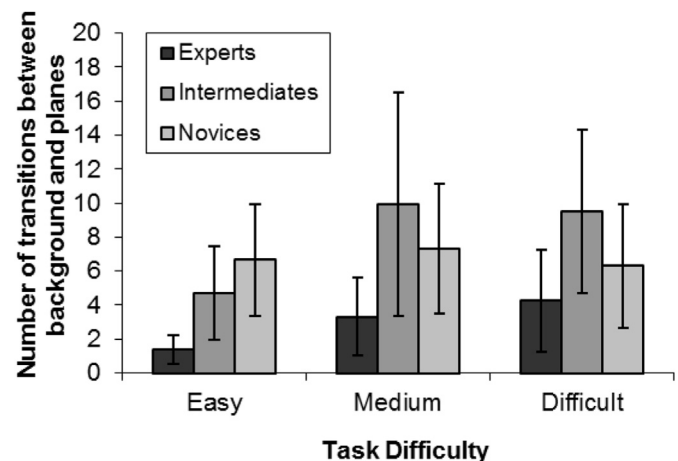


Fig. 5. Means and standard deviations of transitions between the background area and aircraft by experts, intermediates and novices in easy, medium and difficult tasks.

### 3.2. Similarity of task solutions

To test Hypothesis 2 (i.e., similarity) the *similarity of task solutions* was analyzed. Results showed a main effect of expertise level on similarity,  $F(2, 156) = 62.24$ ,  $MSE = .006$ ,  $p < .001$ ,  $\eta_p^2 = .444$ , as well as a main effect of task difficulty,  $F(2, 312) = 870.35$ ,  $MSE = .059$ ,  $p < .001$ ,  $\eta_p^2 = .848$ . Post-hoc tests between expertise levels revealed that experts were more similar to each other in performing the tasks than intermediates ( $p = .003$ ), and both experts and intermediates were more similar to each other than novices (both  $p$ -values  $< .001$ ). Post-hoc tests between task difficulty levels revealed that similarity of task solutions in easy tasks was higher than in medium and difficult tasks, and in medium tasks more similar than in difficult tasks (all  $p$ -values  $< .001$ ). Moreover, results showed an interaction effect between expertise level and task difficulty,  $F(4, 312) = 10.63$ ,  $MSE = .059$ ,  $p < .001$ ,  $\eta_p^2 = .120$  (see Fig. 6). Post-hoc tests revealed that on easy tasks novices were less similar than both intermediates and experts (both  $p$ -values  $< .001$ ). On medium tasks experts were more similar than both intermediates ( $p = .008$ ) and novices ( $p < .001$ ). On difficult tasks novices were less similar than both intermediates ( $p < .001$ ) and experts ( $p = .003$ ).

### 3.3. Performance

To test Hypothesis 3 (i.e., performance), *task correctness score*, *time-on-task* and *perceived mental effort* were analyzed. Results showed a main effect of expertise level on task correctness score,  $F(2, 29) = 13.56$ ,  $MSE = 0.13$ ,  $p < .001$ ,  $\eta_p^2 = .49$ , as well as a main effect of task difficulty,  $F(2, 58) = 27.85$ ,  $MSE = 1.298$ ,  $p < .001$ ,  $\eta_p^2 = .499$ . Post-hoc tests between expertise levels revealed that novices scored significantly lower than both experts ( $p < .001$ ) and intermediates ( $p = .018$ ), while experts and intermediates did not differ significantly from each other ( $p = .187$ ). Post-hoc tests between task difficulty levels revealed that task correctness scores in difficult tasks were lower than in medium tasks and easy tasks (both  $p$ -values  $< .001$ ). No interaction effect between expertise level and task difficulty was found,  $F(4, 58) = 0.788$ ,  $MSE = 1.298$ ,  $p = .533$ ,  $\eta_p^2 = .053$ .

Results showed a main effect of expertise level on time-on-task,  $F(2, 29) = 11.67$ ,  $MSE = 23.12$ ,  $p < .001$ ,  $\eta_p^2 = .455$ , as well as a main effect of task difficulty,  $F(2, 58) = 3.838$ ,  $MSE = 34.029$ ,  $p = .012$ ,  $\eta_p^2 = .215$ . Post-hoc tests between expertise levels revealed that experts performed the tasks significantly faster than both novices ( $p < .001$ ) and intermediates ( $p = .042$ ), while novices and intermediates did not differ significantly from each other ( $p = .184$ ). Post-hoc tests between task difficulty levels revealed that time on task on easy tasks was shorter than on both medium and difficult tasks, and also shorter on medium tasks than on difficult tasks (all

$p$ -values  $< .001$ ). Moreover, results showed an interaction effect between expertise and difficulty level,  $F(4, 58) = 3.84$ ,  $MSE = 34.03$ ,  $p = .012$ ,  $\eta_p^2 = .215$  (see Fig. 7). On easy tasks novices took longer than both experts ( $p < .001$ ) and intermediates ( $p = .042$ ). On medium tasks experts were faster than novices ( $p < .001$ ). On difficult tasks experts were faster than both novices ( $p = .033$ ) and intermediates ( $p < .001$ ).

Finally, results showed a main effect of expertise level on perceived mental effort,  $F(2, 29) = 5.16$ ,  $MSE = .53$ ,  $p = .012$ ,  $\eta_p^2 = .27$ , as well as a main effect of task difficulty,  $F(2, 58) = 150.914$ ,  $MSE = 1.922$ ,  $p < .001$ ,  $\eta_p^2 = .844$ . Post-hoc tests between expertise levels revealed that experts perceived less mental effort in performing the tasks than intermediates ( $p = .011$ ), while, unexpectedly, experts and novices ( $p = .115$ ) as well as intermediates and novices ( $p = .767$ ) did not differ significantly from each other. Post-hoc tests between task difficulty levels revealed that perceived mental effort in easy tasks was lower than in medium and difficult tasks, and in medium tasks was lower than in difficult tasks (all  $p$ -values  $< .001$ ). No interaction effect between expertise level and task difficulty was found,  $F(4, 58) = 1.616$ ,  $MSE = 3.107$ ,  $p = .210$ ,  $\eta_p^2 = .104$ .

## 4. Discussion

The main aim of this study was to investigate expertise differences in visual problem solving strategies and the similarity of task solutions reached at different levels of expertise. Novices, intermediates and experts worked on nine ATC tasks at three levels of difficulty. First, we expected to find differences in eye-movements related to three visual problem solving strategies: Information reduction, perceptual chunking, and means-end analysis. Second, we expected to find differences between expertise levels in the similarity of reached task solutions. Third, we ascertained that performance (i.e., correctness, speed, and perceived mental effort) was indeed different between the three expertise levels. Fourth, we investigated whether task difficulty moderated the effects of expertise.

### 4.1. Visual processes

Our results clearly support the assumption that experts, intermediates and novices apply different visual problem solving strategies (Hypothesis 1). First, there was more information reduction for higher levels of expertise (Hypothesis 1a). Novices

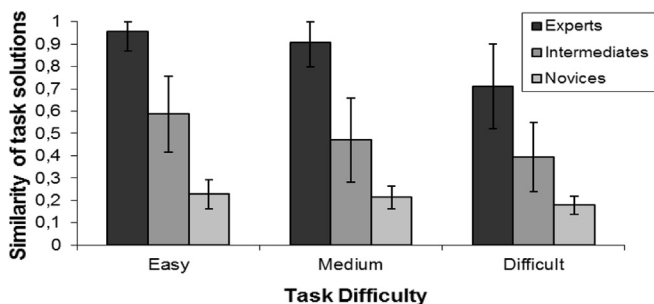


Fig. 6. Means and standard deviations of similarities between the chosen task solutions by experts, intermediates and novices in easy, medium and difficult tasks.

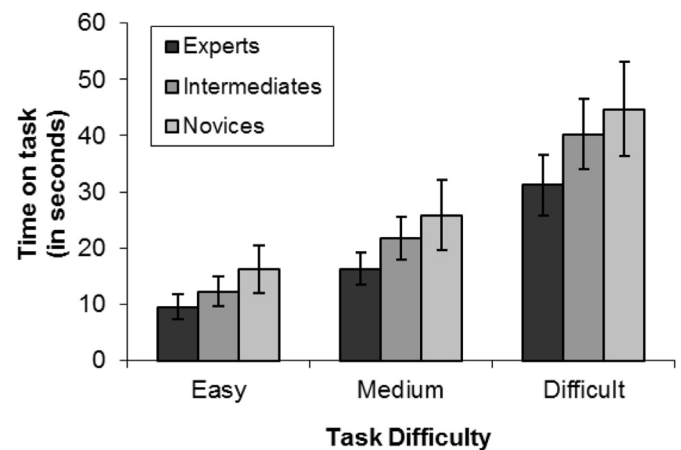


Fig. 7. Means and standard deviations of total time-on-task by experts, intermediates and novices in easy, medium and difficult tasks.

focused much faster, more often, and longer on Artip than intermediates and experts; for intermediates and experts, Artip seems to be an irrelevant area because they simply “know where the goal is”. Furthermore, novices took more time before they first fixated on aircraft. This indicates that they have trouble finding the relevant areas and fixate on irrelevant areas before finding the relevant ones.

Second, experts showed more perceptual chunking of related elements than intermediates and novices (Hypothesis 1b). For experts, fewer fixations on aircraft and fewer transitions between aircraft and the background area were found than for both intermediates and novices. In addition, experts spent less time looking at the background area than intermediates. Thus, experts seem to recognize familiar patterns of task elements (i.e., grouped aircraft) and need to orientate themselves less on the space surrounding these elements to effectively deal with them. These findings clearly demonstrate superior perceptual chunking strategies for experts.

Third, experts seem to use a working-forward rather than means-end strategy (Hypothesis 1c). They showed fewer transitions that included the destination point of the aircraft (i.e., Artip-aircraft and Artip-background) than novices. This indicates that novices frequently focus on the ‘goal’ in order to reach a solution, while experts develop a solution without paying attention to the goal. For intermediates, the number of transitions including Artip is in between that of experts and novices. This suggests a more or less linear development from a means-end strategy to a working-forward strategy.

#### 4.2. Performance

With regard to the similarity of solutions, our results show that solutions of experts are more similar to each other than solutions of intermediates, and that solutions of intermediates are more similar to each other than solutions of novices (Hypothesis 2). Thus, experts recognize a broad range of problem situations which allows them to come up with optimal solutions that are relatively similar across different experts; they all work towards a comparable pre-sorted order of arrival. Intermediates recognize less problem situations and/or have more difficulties to link these situations to their decisions, and novices apply weak problem solving methods such as means-end analysis, both leading to less similarity of their solutions (cf. [Medin et al., 2006](#)). This is an important finding because it suggests that it is worthwhile to teach expert strategies, not only because they are more effective in reaching high performance but also because they are univocal compared to non-expert strategies. Our findings add to earlier findings by [Jarodzka et al. \(2010\)](#) and [Medin, Lynch, Coley, and Atran \(1997\)](#), showing that higher expertise is related to higher similarity in reached solutions.

Experts and intermediates showed better and faster performance than novices (Hypothesis 3). However, the differences between experts and intermediates were not significantly different although in the expected direction. Possibly, the whole set of tasks was not complex enough to reveal performance differences between intermediates, who were nearly certified air traffic controllers, and experts. For perceived mental effort, an unexpected finding is that the effort reported by intermediates is not only higher than that of experts but also than that of novices. On the one hand, this fits the assumption that intermediates may have the knowledge needed to carry out the given tasks but, compared to experts, lag behind in their strategies of efficient information retrieval from the scene (i.e., perceptual chunking). On the other hand, it indicates that novices not only used less effective but also less effort-demanding strategies than intermediates. As a speculation, they experienced

the tasks as so difficult that they were not inclined to invest a high level of mental effort.

Finally, the findings concerning the moderating effects of task difficulty (Hypothesis 4) are equivocal. As expected, the differences between expertise levels are for some of the measures more pronounced for difficult tasks than for easy tasks. For time to first fixation on aircraft, the differences between expertise levels are larger for difficult tasks than for easy tasks; for difficult tasks, expert are quicker than intermediates and intermediates are quicker than novices. For similarity of solutions, similarity decreases somewhat for experts and intermediates as tasks become more complex, but it is consistently low for novices. And for time on task, the difference between novices and experts is smaller for easy tasks than for more difficult tasks. Taken together, these findings suggest that the superior visual problem solving strategies of experts and, to a lesser degree, intermediates yield a greater advantage when working on more difficult tasks.

Yet, some interactions yielded unexpected patterns. Compared to intermediates and experts, novices fixated less on Artip and less on the background area as tasks became more difficult. This indicates that for difficult tasks, novices lose the destination out of sight and seem to become less aware of the whole situation. In other words, the difficult tasks seem to be ‘too difficult’ for the novices, which is also evidenced by their low performance and, as speculated above, their low readiness to invest effort in performing the task. In addition, it is in line with the findings for transitions between background area and aircraft. Experts show more transitions, and thus more awareness of the problem situation as tasks become more difficult; intermediates only show this increase from easy to difficult tasks, and novices do not show such an increase at all.

To summarize, the use of eye-tracking made it possible to unravel the visual problem solving strategies that experts, intermediates and novices use when solving perceptual problems in the complex domain of air traffic control. First, the development of information reduction abilities as described by [Haider and Frensch \(1999\)](#) was demonstrated by faster and longer fixations on relevant areas as expertise increases. Second, the development of schemas that chunk elements together ([Gobet & Simon, 1998](#)) was confirmed by more efficient scan paths as expertise increases. The higher investment of mental effort by intermediates reflects the difficulties they encounter with linking the applicable cognitive schemas to the situation at hand ([Boshuizen & Schmidt, 2008](#)). Third, the change from a means-end approach to a working-forward approach ([Simon, 1975](#)) became evident by focusing less on the final destination point (Artip) as expertise increases. Furthermore, reached solutions became more similar with higher expertise and the more effective strategies of experts often had greater value for more difficult tasks.

#### 4.3. Directions for future research

Our findings indicate two issues that need to be further investigated. First, limited differences in performance were found between intermediates and experts. Future research should contain more complex tasks to create situations in which experts perform significantly higher than intermediates. Such more complex tasks may also be more suitable to unravel their full scale of expertise. Higher levels of complexity could, for example, be obtained by using dynamic traffic situations rather than the stills as used in this study and including modality in the stimuli by presenting more irrelevant information on the screens and/or using audio. Second, future research should aim to explain our unexpected findings for novices, in particular, their low investment of mental effort. It should test our speculation that tasks were experienced as too

difficult by them, leading to a motivational problem and “unwillingness” to invest high effort in solving the visual problems. As self-efficacy is known as a good predictor of performance (Pintrich & De Groot, 1990), self-efficacy measures could be used to provide insight in participants’ belief that one is capable of carrying out a task (Lodewyk & Winne, 2005; Schunk, 1985). Moreover, with better eye-tracking technologies, it will become possible to measure mental effort more directly by the use of eye-tracking (e.g., Klingner, Tversky, & Hanrahan, 2011; May, Kennedy, Williams, Dunlap, & Brannan, 1990).

With regard to theoretical implications, the cognitive theories used to predict our findings turned out to be directly applicable for some visual problem solving processes (e.g., time to first fixation). But for other visual processes (e.g., number of transitions, fixation duration on AOIs) these theories seem to be too limited. More insight in the origin of visual problem solving processes is required and to explain all of our findings, cognitive theories should not only be integrated with visual cognitive (i.e., perceptual) theories but also integrate their description and explanation of the development of attentional processes, chunking processes and problem-solving processes with increasing expertise. With regard to the integration of perceptual aspects, stronger theories should, for example, include the visual integration of information elements in order to explain findings on transitions. With regard to the integration of attentional, chunking and problem-solving processes, such theories should also be able to distinguish more clearly between the working-forward strategies of experts and the strategies of intermediates, which seem to be somewhere between working forward and means-end analysis, in relation to the investment of mental effort and especially the high effort reported by intermediates. As another example, such theories should provide an explanation for experts’ ability to oversee all small but relevant details in visual stimuli and how their strategy to chunk and reduce incoming information elements enables this.

#### 4.4. Practical implications

With regard to practical implications, our findings inform the design of eye-movement modeling examples (EMMEs) that could be used to train visual problem solving strategies for novices, intermediates and experts. Novices use means-end analysis, do not yet have the ability to ignore irrelevant information, and possess no chunks to treat related information elements as one element when solving visual problems. Hence, EMMEs for novices should first show which information is needed to work forward to the goal instead of backward from the goal. That way, the learner is shown how decisions are made without taking the general destination point into account. Second, they must indicate which information is relevant for problem solving, and where this information is located in the complex visual representation. Third, they should make visible which related information elements can be treated as chunks. And finally, they should reflect the divergence in visual problem solving strategies applied by experts (Van Merriënboer & Kirschner, 2013) because the similarity of expert strategies is yet relatively high but might still lead to slightly different solutions.

For intermediates, EMMEs should primarily take into account their tendency to focus on irrelevant information resulting in a relatively high cognitive load. They should help them to reduce visual search by focusing on the information that is minimally required to take safe decisions. For example, EMMEs could be based on prototypical situations for which it is known that visual problem solving profits from the use of a perceptual chunking strategy (e.g., Gobet & Simon, 1998). Then, intermediates could learn how to recognize the most relevant information from a related group of

objects and, next, why the grouped elements are crucial to rely on in a certain situation (i.e., perceptual chunking).

For experts, EMMEs could be used to train them in working with newly introduced technologies or regulations, which might require the observation of new information elements during visual problem solving (e.g., additional information in aircraft labels from new spacing tools). In such situations, EMMEs could be helpful in the same way as they are for intermediates. Furthermore, EMMEs based on the eye-movements of peers or on own eye-movements could help to foster reflection on the use of own visual problem solving strategies and so contribute to a process of deliberate practice (Ericsson, 2004).

A limitation of our study is that solely eye-tracking measures were used to reveal visual problem solving strategies. A triangulation of data including – in addition to eye-tracking measures – self-reports (e.g., cued retrospective reports) and/or questionnaires could further disclose the use of visual problem solving strategies and unravel their relation to, for example, knowledge structures and motivation. Furthermore, the number of participants in our study was relatively low because there are not many experts in ATC available, participation of these experts is expensive, and eye-tracking data require much time and effort to analyze. Although the number of participants per condition in our study is comparable to that of similar expertise research (cf. Gegenfurtner et al., 2011), results could have been more pronounced with more participants.

Concluding, this study gave insight in three developmental phases of visual problem solving strategies across a range of task difficulties. It showed that strategies are clearly different for different levels of expertise and lead to more similar solutions as expertise increases. The findings provide important implications for the design of EMMEs for training in complex visual domains. Our study showed that care must be taken in selecting eye-movement models in order to fit the development level of learners’ cognitive schemas and their related visual problem solving strategies.

#### Acknowledgment

This study was part of the future air-traffic management program funded by HYPERLINK “gs1:” \o “gs1:” Knowledge Development Centre Mainport Schiphol. We would like to thank prof. dr. H.P.A. Boshuizen for her useful comments on an earlier version of this manuscript.

#### References

- Bellenkes, A. H., Wickens, C. D., & Kramer, A. F. (1997). Visual scanning and pilot expertise: the role of attentional flexibility and mental model development. *Aviation Space and Environmental Medicine*, 68, 569–579.
- Berliner, D. C. (1986). In pursuit of the expert pedagogue. *Educational Researcher*, 15(7), 5–13. <http://dx.doi.org/10.3102/0013189X015007007>.
- Boshuizen, H. P. A., & Schmidt, H. G. (1992). On the role of biomedical knowledge in clinical reasoning by experts, intermediates and novices. *Cognitive Science*, 16, 153–184. [http://dx.doi.org/10.1207/s15516709cog1602\\_1](http://dx.doi.org/10.1207/s15516709cog1602_1).
- Boshuizen, H. P. A., & Schmidt, H. G. (2008). The development of clinical reasoning expertise. In J. Higgs, M. A. Jones, S. Loftus, & N. Christensen (Eds.), *Clinical reasoning in the health professions* (3rd ed.) (pp. 113–122). Oxford, UK: Butterworth Heinemann, Elsevier.
- Chi, M. T. H., Glaser, R., & Rees, E. (1982). *Expertise in problem solving*. In R. J. Sternberg (Ed.), *Advances in the psychology of human intelligence* (Vol. 1); (pp. 7–75). Hillsdale, NJ: Erlbaum.
- De Groot, A. D. (1978). *Thought and choice in chess* (2nd ed.). The Hague, The Netherlands: Mouton.
- Dreyfus, H. L., & Dreyfus, E. D. (2005). Peripheral vision: expertise in real world contexts. *Organization Studies*, 26, 779–792. <http://dx.doi.org/10.1177/0170840605053102>.
- Endsley, M. R. (1995). Toward a theory of situation awareness in dynamic systems. *Human Factors*, 37, 32–64. <http://dx.doi.org/10.1518/001872095779049543>.
- Endsley, M. R. (2006). Expertise and situation awareness. In K. A. Ericsson, N. Charness, R. R. Hoffman, & P. J. Feltovich (Eds.), *The Cambridge handbook of expertise and expert performance* (pp. 633–652). New York: Cambridge University Press.



- Ericsson, K. A. (2006). The influence of experience and deliberate practice on the development of superior expert performance. In K. A. Ericsson, N. Charness, R. R. Hoffman, & P. J. Feltovich (Eds.), *The Cambridge handbook of expertise and expert performance* (pp. 685–706). New York: Cambridge University Press.
- Ericsson, K. A., & Lehmann, A. C. (1996). Expert and exceptional performance: evidence of maximal adaptation to task constraints. *Annual Review of Psychology*, 47, 273–305. <http://dx.doi.org/10.1146/annurev.psych.47.1.273>.
- Feldon, D. F. (2007). The implications of research on expertise for curriculum and pedagogy. *Educational Psychology Review*, 19, 91–110. <http://dx.doi.org/10.1007/s10648-006-9009-0>.
- Gegenfurtner, A., Lehtinen, E., & Säljö, R. (2011). Expertise differences in the comprehension of visualizations: a meta-analysis of eye-tracking research in professional domains. *Educational Psychology Review*, 23, 523–552. <http://dx.doi.org/10.1007/s10648-011-9174-7>.
- Gegenfurtner, A., Siewiorek, A., Lehtinen, E., & Säljö, R. (2013). Assessing the quality of expertise differences in the comprehension of medical visualizations. *Vocations and Learning*, 6, 37–54. <http://dx.doi.org/10.1007/s12186-012-9088-7>.
- Gobet, F., & Simon, H. A. (1998). Expert chess memory: revisiting the chunking hypothesis. *Memory*, 6, 225–255. <http://dx.doi.org/10.1080/741942359>.
- Goldstone, R. L. (1998). Perceptual learning. *Annual Review of Psychology*, 49, 585–612. <http://dx.doi.org/10.1146/annurev.psych.49.1.585>.
- Gronlund, S. D., Dougherty, M. R. P., Durso, F. T., Canning, J. M., & Mills, S. H. (2005). Planning in air traffic control: impact of problem type. *International Journal of Aviation Psychology*, 15, 269–293. [http://dx.doi.org/10.1207/s15327108ijap1503\\_4](http://dx.doi.org/10.1207/s15327108ijap1503_4).
- Haider, H., & Frensch, P. A. (1999). Eye movement during skill acquisition: more evidence for the information-reduction hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 25, 172–190. <http://dx.doi.org/10.1037/0278-7393.25.1.172>.
- Hoffman, R. R. (1987). The problem of extracting the knowledge of experts from the perspective of experimental psychology. *AI Magazine*, 8(2), 53–67. Retrieved from <http://www.aaai.org>.
- Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., & Van de Weijer, J. (2011). *Eye tracking: A comprehensive guide to methods and measures*. Oxford, UK: Oxford University Press.
- Jarodzka, H., Balslev, T., Holmqvist, K., Nyström, M., Scheiter, K., Gerjets, P., et al. (2012). Conveying clinical reasoning based on visual observation via eye-movement modelling examples. *Instructional Science*, 40, 813–827. <http://dx.doi.org/10.1007/s11251-012-9218-5>.
- Jarodzka, H., Scheiter, K., Gerjets, P., & Van Gog, T. (2010). In the eyes of the beholder: how experts and novices interpret dynamic stimuli. *Learning and Instruction*, 20, 146–154. <http://dx.doi.org/10.1016/j.learninstruc.2009.02.019>.
- Jarodzka, H., Van Gog, T., Dorr, M., Scheiter, K., & Gerjets, P. (2013). Learning to see: guiding students' attention via a model's eye movements fosters learning. *Learning and Instruction*, 25, 62–70. <http://dx.doi.org/10.1016/j.learninstruc.2012.11.004>.
- Jongman, R. W. (1968). *Het oog van de meester [The eye of the master]*. Assen, The Netherlands: Van Gorcum & Comp. n.v.
- Kasarskis, P., Stehwi, J., Kickox, J., & Aretz, A. (2001). Comparison of expert and novice scan behaviors during VFR flight. In *Paper presented at the 11th international symposium on aviation psychology*. Columbus, OH.
- Klingner, J., Tversky, B., & Hanrahan, P. (2011). Effects of visual and verbal presentation on cognitive load in vigilance, memory, and arithmetic tasks. *Psychophysiology*, 48, 323–332. <http://dx.doi.org/10.1111/j.1469-8986.2010.01069.x>.
- Lesgold, A., Rubinson, H., Feltovich, P., Glaser, R., Klopfer, D., & Wang, Y. (1988). Expertise in a complex skill: diagnosing x-ray pictures. In M. T. H. Chi, R. Glaser, & R. J. Farr (Eds.), *The nature of expertise* (pp. 311–342). Hillsdale, NJ: Erlbaum.
- Levenshtein, V. (1966). Binary codes capable of correcting deletions, insertions and reversals. *Soviet Physics – Doklady*, 10, 707–710. Retrieved from <http://www.scribd.com/>.
- Lodewyk, K. R., & Winne, P. H. (2005). Relations among the structure of learning tasks, achievement, and changes in self-efficacy in secondary students. *Journal of Educational Psychology*, 97, 3–12. <http://dx.doi.org/10.1037/0022-0663.97.1.3>.
- Lowe, R. K. (2003). Animation and learning: selective processing of information in dynamic graphics. *Learning and Instruction*, 13, 157–176. [http://dx.doi.org/10.1016/S0959-4752\(02\)00018-X](http://dx.doi.org/10.1016/S0959-4752(02)00018-X).
- May, J. G., Kennedy, R. S., Williams, M. C., Dunlap, W. P., & Brannan, J. R. (1990). Eye movement indices of mental workload. *Acta Psychologica*, 75(1), 75–89. [http://dx.doi.org/10.1016/0001-6918\(90\)90067-P](http://dx.doi.org/10.1016/0001-6918(90)90067-P).
- Mayer, R. E. (2005). Cognitive theory of multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 31–48). New York: Cambridge University Press.
- Mayer, R. E., & Moreno, R. (2003). Nine ways to reduce cognitive load in multimedia learning. *Educational Psychologist*, 38, 43–52. [http://dx.doi.org/10.1207/s15326985ep3801\\_6](http://dx.doi.org/10.1207/s15326985ep3801_6).
- Medin, D. L., Lynch, E. B., Coley, J. D., & Atran, S. (1997). Categorization and reasoning among tree experts: do all roads lead to Rome? *Cognitive Psychology*, 32, 49–96. <http://dx.doi.org/10.1006/cogp.1997.0645>.
- Medin, D. L., Ross, N. O., Atran, S., Cox, D., Coley, J., Proffitt, J. B., et al. (2006). Folkbiology of freshwater fish. *Cognition*, 99, 237–273. <http://dx.doi.org/10.1016/j.cognition.2003.12.005>.
- Mumford, M. D., Schultz, R. A., & Van Doorn, J. R. (2001). Performance in planning: processes, requirements, and errors. *Review of General Psychology*, 5, 213–240. <http://dx.doi.org/10.1037/1089-2680.5.3.213>.
- Oprins, E., & Schuwer, M. (2003). Competentiegericht opleiden en beoordelen bij LVNL [Competence-based training and assessment at LVNL]. *HUFAG Nieuwsbrief*, 6, 2–4. Retrieved from <http://www.hufag.nl>.
- Paas, F. (1992). Training strategies for attaining transfer of problem-solving skill in statistics: a cognitive-load approach. *Journal of Educational Psychology*, 84, 429–434. <http://dx.doi.org/10.1037/0022-0663.84.4.429>.
- Pintrich, P. R., & De Groot, E. V. (1990). Motivational and self-regulated learning components of classroom academic performance. *Journal of Educational Psychology*, 82, 33–40.
- Reingold, E. M., & Sheridan, H. (2011). Eye movements and visual expertise in chess and medicine. In S. P. Livensedge, I. D. Gilchrist, & S. Everding (Eds.), *The Oxford handbook of eye movements* (pp. 523–562). Oxford, UK: Oxford University Press.
- Scheiter, K., Gerjets, P., Huk, T., Imhof, B., & Kammerer, Y. (2009). The effects of realism in learning with dynamic visualizations. *Learning and Instruction*, 19, 481–494. <http://dx.doi.org/10.1016/j.learninstruc.2008.08.001>.
- Schmidt, R. G., Norman, G. R., & Boshuizen, H. P. A. (1990). A cognitive perspective on medical expertise: theory and implications. *Academic Medicine*, 65, 611–621. Retrieved from <http://journals.lww.com/academicmedicine>.
- Schunk, D. H. (1985). Self-efficacy and school learning. *Psychology in the Schools*, 22, 208–223.
- Simon, H. A. (1975). The functional equivalence of problem solving skills. *Cognitive Psychology*, 7, 268–288. [http://dx.doi.org/10.1016/0010-0285\(75\)90012-2](http://dx.doi.org/10.1016/0010-0285(75)90012-2).
- Spanjers, I., Van Gog, T., & Van Merriënboer, J. J. G. (2010). A theoretical analysis of how segmentation of dynamic visualizations optimizes students' learning. *Educational Psychology Review*, 22, 411–423. <http://dx.doi.org/10.1007/s10648-010-9135-6>.
- Spivey, M. J., & Dale, R. (2011). Eye movements both reveal and influence problem solving. In S. P. Livensedge, I. D. Gilchrist, & S. Everding (Eds.), *The Oxford handbook of eye movements* (pp. 551–562). Oxford, UK: Oxford University Press.
- Sweller, J. (2004). Instructional design consequences of an analogy between evolution by natural selection and human cognitive architecture. *Instructional Science*, 32, 9–31. <http://dx.doi.org/10.1023/B:TRUC.0000021808.72598.4d>.
- Sweller, J., Van Merriënboer, J. J. G., & Paas, F. (1998). Cognitive architecture and instructional design. *Educational Psychology Review*, 10, 251–296. <http://dx.doi.org/10.1023/A:1022193728205>.
- Van Gog, T., Jarodzka, H., Scheiter, K., Gerjets, P., & Paas, F. (2009). Attention guidance during example study via the model's eye movements. *Computers in Human Behavior*, 25, 785–791. <http://dx.doi.org/10.1016/j.chb.2009.02.007>.
- Van Gog, T., Paas, F., & Van Merriënboer, J. J. G. (2006). Effects of process-oriented worked examples on troubleshooting transfer performance. *Learning and Instruction*, 16, 154–164. <http://dx.doi.org/10.1016/j.learninstruc.2006.02.003>.
- Van Gog, T., Paas, F., & Van Merriënboer, J. J. G. (2008). Effects of studying sequences of process-oriented and product-oriented worked examples on troubleshooting transfer efficiency. *Learning and Instruction*, 18, 211–222. <http://dx.doi.org/10.1016/j.learninstruc.2007.03.003>.
- Van Gog, T., & Rummel, N. (2010). Example-based learning: integrating cognitive and social-cognitive research perspectives. *Educational Psychology Review*, 22, 155–174. <http://dx.doi.org/10.1007/s10648-010-9134-7>.
- Van Merriënboer, J. J. G., Clark, R., & Croock, M. B. M. (2002). Blueprints for complex learning: the 4C/ID-model. *Educational Technology Research and Development*, 50, 39–61. <http://dx.doi.org/10.1007/bf02504993>.
- Van Merriënboer, J. J. G., & Kirschner, P. A. (2013). *Ten steps to complex learning* (2nd Rev. ed.). New York: Routledge.